



Backend.AI:

왜 우리는 우리 핵심 제품을 오픈소스화 했는가

Lablup Inc.

2017. 11. 20



Backend.AI 프로젝트 소개	01
Backend.AI 데모	02
Backend.AI 오픈소스 프로젝트	03
오픈소스와 래블업	04
마치며	05

Backend.AI 프로젝트 소개

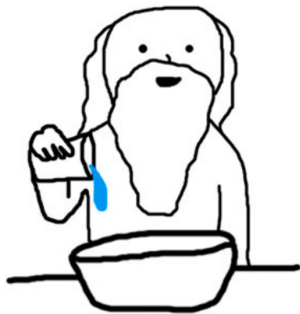




- 오픈소스와 클라우드의 발달 – 정말로 AI 하기가 쉬워졌을까?

신이 AI 분야를 만들 때

일단 오픈소스
한컵 붓고..



클라우드 살짝 엮고..



딥러닝을 잘 부으면...
다 되는 건가?



(사실은 삽질 삽질 삽질
삽질 삽질을... 으어어)



<http://kr.vonvon.me/quiz/329>,
<http://kr.vonvon.me/quiz/1633>



- 팩트!

- 오픈소스 생태계와 이를 활용한 클라우드 인프라가 점점 빠르게 좋아지고 있다
- Amazon Glue / GluOn
- Azure ML Workbench / Studio
- Google Cloud Datalab / DialogFlow
- Naver nsml / C3 DL

- 문제점

- AI 기반 기술의 발전 속도 >>> 연구·개발자들의 AI 활용능력 발전 속도 >>>> 인간사회의 AI 대응 속도
- 플랫폼이 아무리 좋아져도 **data locality 문제**와 **규제**는 근본적으로 피해가기 어렵다
- 플랫폼이 아무리 쉬워져도 여전히 **비개발자**에겐 어렵다



쉽게

- 원래 사용하던 방식과 같도록
- 최소한의 환경설정만으로

빠르게

- GPU를 잘 활용할 수 있도록
- 원하는 만큼 연산자원을 바로바로

값싸게

- 사용한 만큼만 지불하도록
- 비용 제한에 맞춰 성능 조정을 자동으로

함께

- 한번 만들면 누구나 똑같이 재현하고 재사용하도록
- 다른 사람들과 충돌 없이 자원을 공유할 수 있도록

어디서나

- 내가 가진 서버를 지금 바로 활용하거나
- 이도저도 귀찮다면 클라우드에 맡겨서



Cloud

Pay-as-you-go

계산 기반의 과학·공학
연구, 딥러닝 모델 훈련
및 엄청나게 간편한 코딩
교육 환경을 구축하는
Scalable PaaS

Ground

Bring Your
Own Hardware

자신만의 Lablup.AI
서버팜을 설치하고
수정하고 개발할 수 있는
오픈소스 버전

Garden

Showcases

Lablup.AI 사용자들을
위한 문서, 포럼 및
쇼케이스들

Backend.AI

codeonweb.com



쉽게

- Jupyter, VS Code, Atom 플러그인 제공
- API key만 설정하면 끝!

빠르게

- 컨테이너와 GPU 기술을 결합
- 1초 이내에 연산 세션이 뜨도록

값싸게

- 밀리초/KiB 단위까지 정밀한 자원 사용 측정
- (향후 지원 예정!)

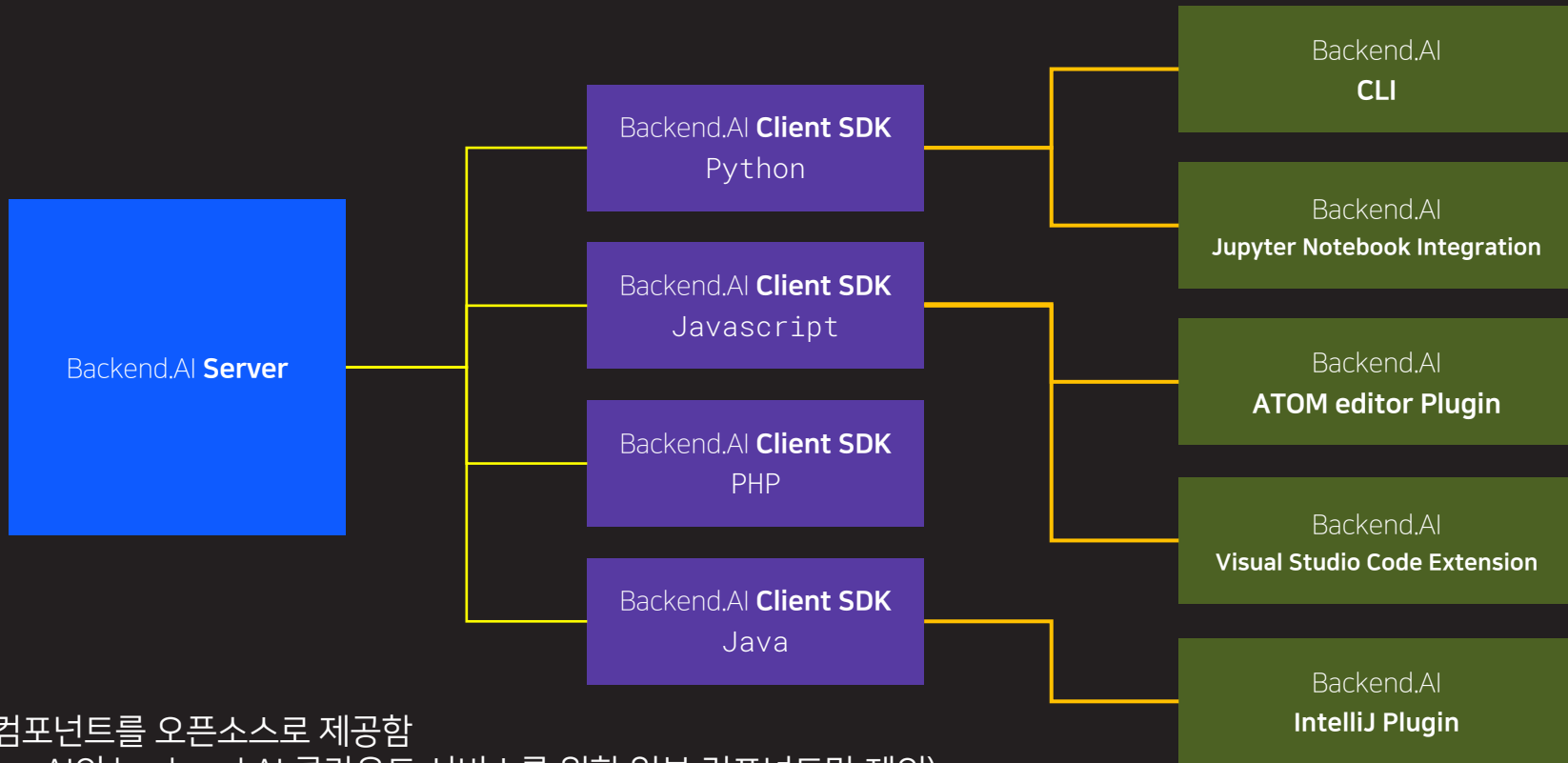
함께

- 컨테이너를 활용한 언어별·버전별 가상 환경 제공
- 시스템콜 샌드박싱 + Docker의 자원제한 고도화

어디서나

- 오픈소스 버전 제공 (github.com/lablup/backend.ai)
- 클라우드 버전 제공 (cloud.backend.ai)

Backend.AI 구조



모든 컴포넌트를 오픈소스로 제공함
(Lablup.AI의 backend.AI 클라우드 서비스를 위한 일부 컴포넌트만 제외)



- 빠른 시작
 - 클라우드 가입만으로 바로 사용 가능
 - 사용자의 요청 즉시 가상 프로그래밍 환경 생성
- 다양한 요구사항 충족
 - 모든 주요 프로그래밍 언어와 런타임 지원
Python, R, Julia, Octave, PHP, Go, C/C++, Java, NodeJS, Lua, Haskell, Rust
 - 동일한 기계학습 라이브러리의 여러 버전 동시 지원
TensorFlow, Caffe, PyTorch, Keras
- 친숙한 사용자 경험 + 개발자 친화적 프레임워크
 - 기존 연구·개발자들에게 익숙한 환경과의 통합 지원 (코드편집기, 웹기반 연구노트)
Jupyter, VS Code, Atom Editor, IntelliJ^{beta}
 - `$ backend.ai run` 명령줄 및 클라우드 인터프리터·컴파일러 지원
 - 개발자를 위한 HTTP 기반 공용 API 및 언어별 SDK 제공
Python, Javascript, PHP^{beta}



- 고성능
 - 비동기 I/O 기반 저지연 컨테이너 풀링 기법 (Python asyncio + Docker)
 - 컨테이너에서의 머신러닝 GPU 가속 지원 (nvidia-docker)
- 다중 사용자 보안
 - 다이나믹 샌드박싱 : 가상환경별로 특화된 시스템콜 제어 및 필터링
 - 레거시 소프트웨어를 위한 컨테이너 리소스 제한 강제 기술
- 하이퍼 스케일링 Hyper-scaling
 - On-premise 단독 private cluster
 - Hybrid cloud (on-premise + public cloud)
 - Public cloud (AWS + MS Azure + Google Cloud 조합)
 - 계산종류·부하·설정에 따라 특화된 클라우드로 사용자 요청 routing

Language & Frameworks

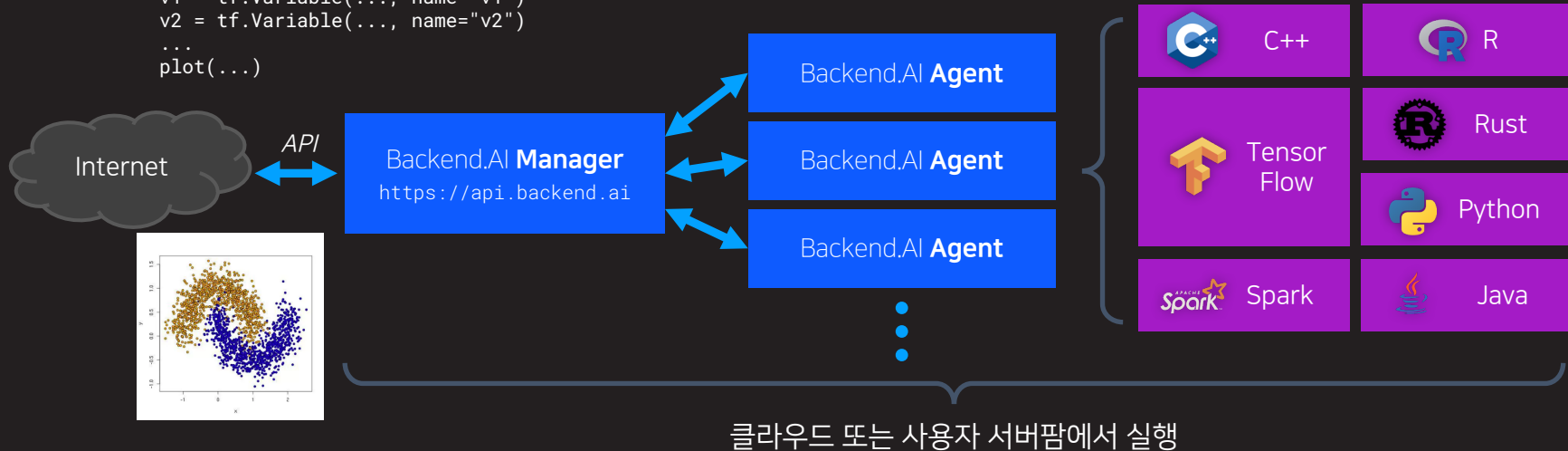


System	Languages	Frameworks	Developed Framework
Backend.AI Ground	Python 3.6	aiodocker aiozmq alembic pyzmq SQLAlchemy nvidia-docker	Backend.AI-manager Backend.AI-agent
	Go		Backend.AI-Jail
	ETC.	Docker 17.9	
Backend.AI Cloud CodeOnWeb.com	Python 3.6	Django 1.11	Customized Django Site-independent overloading
	Javascript ES6 / HTML5	Polymer 2.2	Lablup-webcomponents

Backend.AI Server 구조



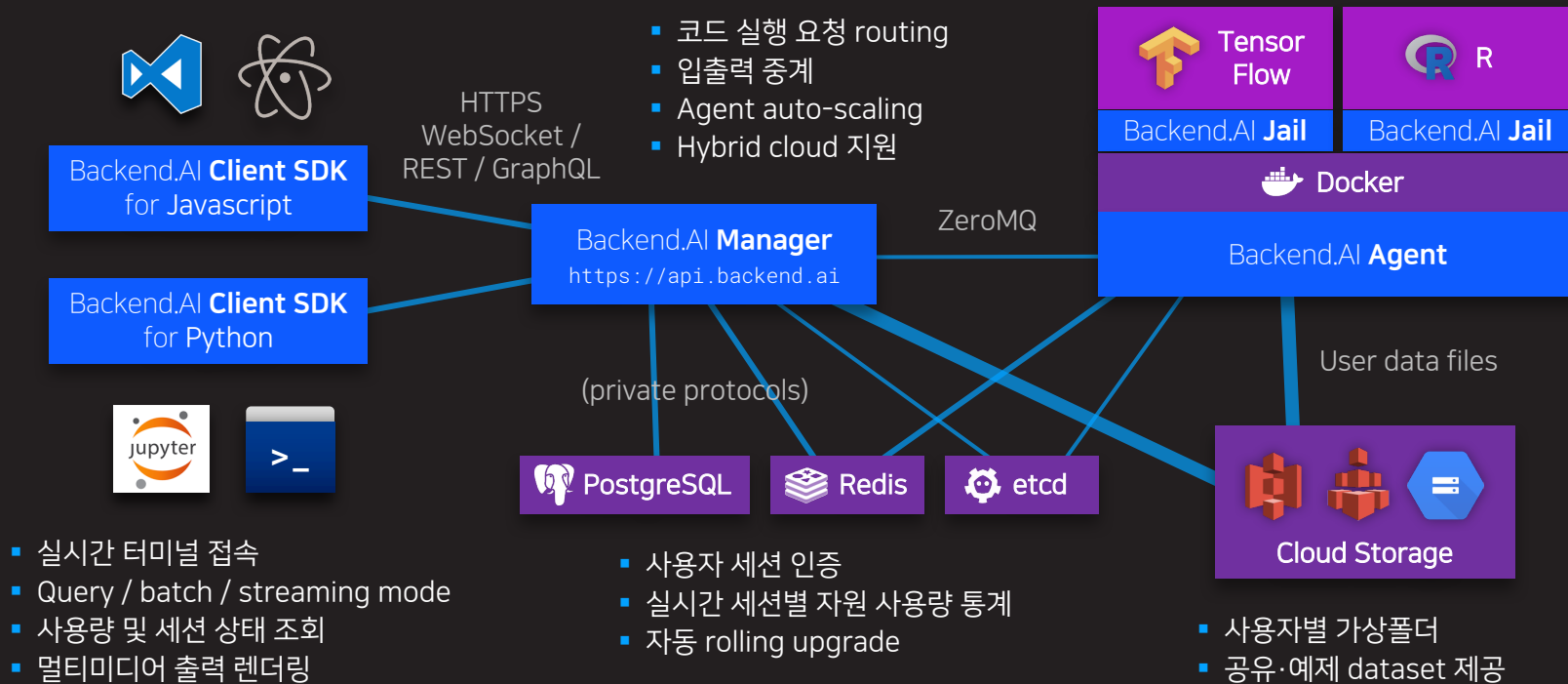
```
import tensorflow as tf
import matplotlib
v1 = tf.Variable(..., name="v1")
v2 = tf.Variable(..., name="v2")
...
plot(...)
```



Backend.AI 상세 구조



- 프로그래밍할 수 있는 시스템콜 샌드박스
- 코어 개수, 메모리 등 컨테이너 자원 제한



Backend.AI 데모

백문이 불여일견!





EXPLORER

OPEN EDITORS

start.md

LIVE-CODE-RUNNER-EXA...

julia.jl

lua.lua

JS nodejs.js

octave.m

php7.php

Python DL Examples...

python3.py

r.r

rust.rs

shellscript.sh

start.md

start2.md

swift.swift

tensorflow_mnist_de...

tensorflow.py

test.py

theano.py

torch.py

Untitled.ipynb

Untitled1.ipynb

Untitled2.ipynb

Untitled3.ipynb

start.md x

```
1 live-code-runner
2 =====
3
4 * Let's start!
5 * You can install the package via `preferences -> install` and search
6   `live-code-runner`
7 * All you need is `API key` from https://cloud.backend.ai
8
9 # How to run?
10 * You can either run via `context menu` or `command palette.`
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

Backend.AI

[LOG] Running...

[LOG] Finished. (33.183 sec.)



```
1. inureyes@Cruise-LAN: ~/Development/backend.ai/jupyter-demo (zsh)
../jupyter-demo (zsh) %1
(env) → jupyter-demo _
```

GitHub, Inc. requests Issues Marketplace Explore

to innovate research / education processes.
blup.com contact@lablup.com

Projects 1 Settings

end.ai-client-py
Talkativot: Do-It-Yourself
friend
Python ★ 5 🍴 3

Type: All Language: All

Top languages
Python HTML
TypeScript Ju

Most used topics
sorna python
cloud-computing
asyncio

 Summary

 Entries

 **Jobs**

 Statistics

 Settings TEST

 Billing

 Staff

[Report](#) · [Terms of Service](#) · [Privacy Policy](#)

Works

Running

#	Job ID	Starts	Configuration		Using		Control					
0	APDKQMFNA	May 16 11:40:12 (52sec.)	1core	2GB	2GB	121sec.	551MB					
1	APDWKDJNFI	May 14 22:23:02 (~2d.)	8core	96GB	128GB	339hr.	1362MB					



 Summary

 Entries

 Jobs

 Statistics

 **Settings** **TEST**

 Billing

 Staff

[Report](#) · [Terms of Service](#) ·
[Privacy Policy](#)

Computation

General

Current storage 0 GB
User library and files

Preserve results in Jobs ☒
Every results will be recorded

Cost estimation 0.0\$
Estimated cost by settings above **Free Tier**

Machine Learning

Concurrent instances 
Size of instance bundle

Memory per instance  16 GB
Memory for one instance



Setting

FOR TESTERS ONLY

Jeongkyu Shin



Cost estimation

Estimated cost by settings above

0.0\$

Free Tier

Machine Learning

Concurrent instances

Size of instance bundle



Memory per instance

Memory for one instance



GB

Cost

Cost limit for a day

Maintain workload in selected cost



\$

Automatic cost optimization

Let Sorna optimize everything



Backend.AI 오픈소스 이야기

우리는 어쩌다 오픈소스 플랫폼을 만들게 되었나







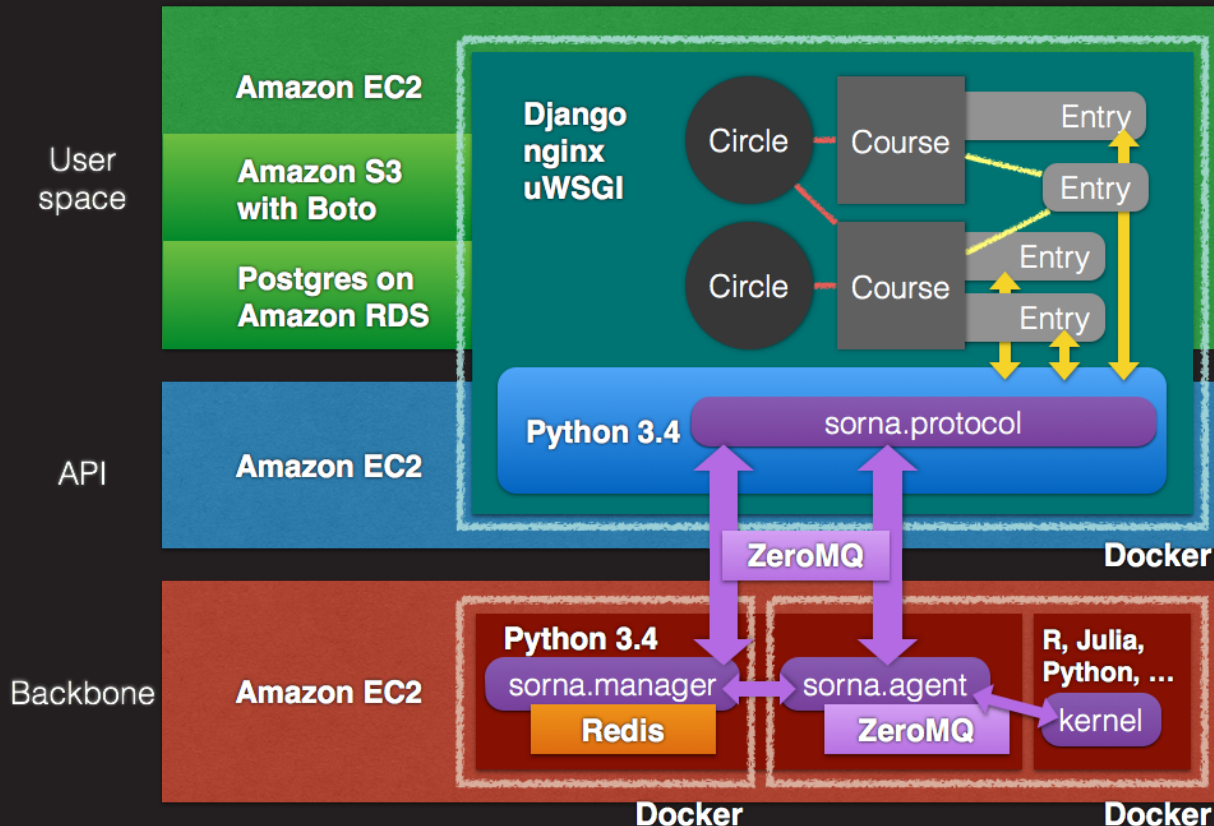
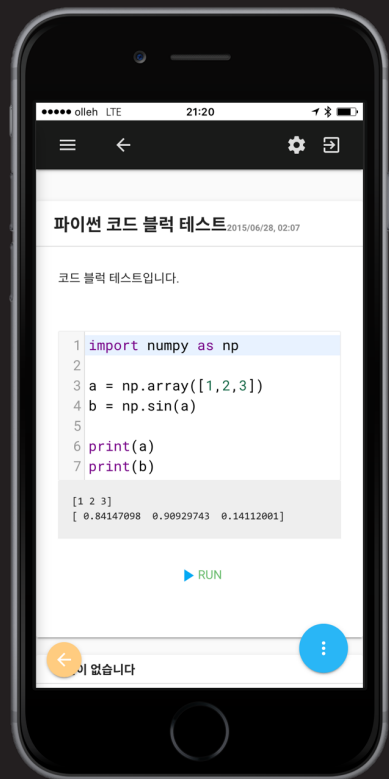
연구말싸미 등귀에 달아 코드로 서르 사맛띠 아니할싸
이런 전차로 어린 학도가 니르고져 홀빼이셔도 마참내
제 뜨들 시러퍼디 못할 노미 하니라.

내 이랄 위하야 어엿비 너겨 새로 **연구 코드 나누미**와
과학 공학 노리터를 망가노니 사람마다 희여 수비니겨
날로 쑤메 뻘한키 하고져 할따라미니라.

2015년에 만들려고 보니까 이걸 받쳐줄 백엔드가 필요하다!
지금이야 좋은 오픈소스가 많지만 그때만 해도... ㅠㅠ



Backend.AI 개발 스토리 - 2015년





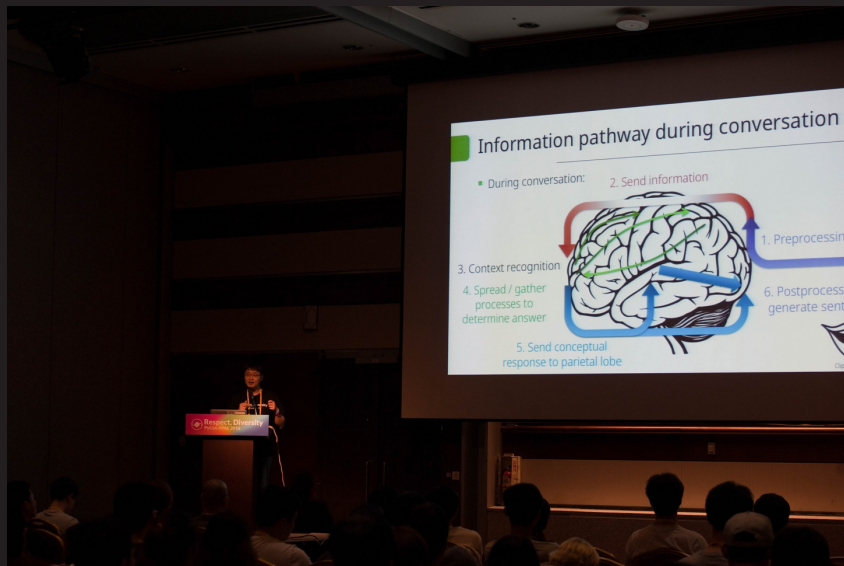
- 2015년 중순
 - 누구도 놀라울 정도로 머신 러닝에 관심이 없었음
 - 플랫폼 설명하러 가면: 머신러닝을 사용할 곳이 있겠어요? 학생들 시키면 돈 안내도 해 오는데요?
 - 간단한 산수 후...
- 북극에서 냉장고를 파는게 진정한 사업가라던데
 - 우린 그냥 아프리카에 냉장고 파는게 낫겠다
- 오픈소스!
 - 구성원들의 거부감이 없음
 - 왜? 오픈소스하다가 만난 사람들끼리 창업을 했음.
- 시점
 - 코드 공개해도 안 쪽팔릴 때

Backend.AI 개발 스토리 - 2016년



<http://www.hani.co.kr/arti/sports/baduk/734599.html>

Backend.AI 개발 스토리 - 2016년



<https://www.flickr.com/photos/126829363@N08/29442206464/in/album-72157671256038363/>

Building AI Chat bot using Python 3 & TensorFlow



Jeongkyu Shin
Lablup Inc.



Illustration * © Idol M@ster / Bandal Namco Games. All rights reserved.



feat. 이말년씨리즈



- 오픈소스화 및 github 저장소 공개: SOSCON 2016



lablup / backend.ai-manager

<> Code

Issues 19

Pull requests 0

Projects 4

Wiki

In

refs #8, #19, #28, #29, #31, #32, #33: Rewrite the manager core.

- * Now manager + gateway runs on multiple CPU cores with sane transaction semantics. Thanks to aiotoools!
 - It no longer depends on Redis as a pub-sub broker nor a database. All communications are done via ZeroMQ with no centralized queue server.
 - Redis is used only for per-keypair rate-limiting.
- * Now the manager searches available agent to spawn new containers based on available memory / CPU / GPU capacity units. No more hard-coded instance types!
- * The db schema is now prepared for multi-container kernel sessions.
 - User-facing APIs now use "session ID" which is directed to the master container of the given session.
 - Each container has unique "kernel ID" and managed individually.
- * Replace asyncpg + asyncpgsa with aiopg for better SQLAlchemy supports (especially custom type decorators).

- Sorna → Backend.AI 브랜드 변경
- 데이터 분석 및 딥러닝을 위한 안정성 강화
- 단일 연산 세션에 대한 다중 컨테이너 지원 준비
- GPU 고려한 스케줄링 강화
- Python 3.6 + aiotoools 기반
- aiocoder 프로젝트의 정식 aio-libs 승격
- 명령줄 인터페이스 편의성 개선
- 단일 PC 설치 가능하도록 매뉴얼·코드 정비
- ...
- v1.0 release!!



- Team communication
 - Slack → Synology Chat (+Synology CloudStation) → Microsoft Teams (+OneNote+OneDrive)
- Issue tracker
 - GitHub + Microsoft Teams
- Autodeploy
 - Lablup.AI : GitHub + post-commit signal to Azure → Autobuild on Azure webapp
 - CodeOnWeb: manual via script execution (to keep users safe)
- CI
 - Travis CI + Datadog
- Error reporting
 - Sentry + Datadog



This organization

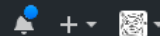
backend.ai

Pull requests

Issues

Marketplace

Explore



Lablup

lab | up: Make AI Accessible - A start-up to innovate research / education processes.

📍 Seoul, Republic of Korea

🌐 <http://www.lablup.com>✉ contact@lablup.com

📁 Repositories 68

👤 People 8

👥 Teams 4

📁 Projects 1

⚙ Settings

Pinned repositories

Customize pinned repositories

≡ backend.ai

Meta-repository for backend.ai

● Python ★ 34 🍴 5

≡ backend.ai-client-py

Client API Library for Lablup Backend.AI Service

● Python ★ 4 🍴 1

≡ talkativot

Talkativot: Do-It-Yourself backbone for your AI friend

● Python ★ 5 🍴 3

backend.ai

Type: All ▾

Language: All ▾

📁 New

16 results for repositories matching backend.ai

✕ Clear filter

backend.ai

Meta-repository for backend.ai

[python](#) [api](#) [docker](#) [documentation](#) [distributed-computing](#)

● Python ★ 34 🍴 5 📄 LGPL-3.0 Updated 2 days ago

Top languages

● Python ● HTML ● JavaScript
● TypeScript ● C

Most used topics

Manage

[sorna](#) [python](#)

PACKAGE INDEX >>

[Browse packages](#)
[List trove classifiers](#)
[RSS \(latest 40 updates\)](#)
[RSS \(newest 40 packages\)](#)
[Terms of Service](#)
[PyPI Tutorial](#)
[PyPI Security](#)
[PyPI Support](#)
[PyPI Bug Reports](#)
[PyPI Discussion](#)
[PyPI Developer Info](#)

ABOUT >>

NEWS >>

DOCUMENTATION >>

DOWNLOAD >>

COMMUNITY >>

FOUNDATION >>

CORE DEVELOPMENT >>

backend.ai 1.0.2

Lablup Backend.AI Meta-package

[Downloads ↓](#)

Backend.AI is a streamlined backend service framework hosting heterogeneous programming languages and popular AI frameworks. It manages the underlying computing resources for multi-tenant computation sessions where such sessions are spawned and executed instantly on demand.

All sub-projects are licensed under LGPLv3+.

By installing this meta-package, you get the client with command-line interface by default and optionally you may add the manager and agent using pip extra tags.

```
$ pip install backend.ai
(installs the common and client libs which includes CLI)
$ pip install backend.ai[manager]
(installs the common and client libs with the manager/gateway daemon)
$ pip install backend.ai[agent]
(installs the common and client libs with the agent daemon)
```

Server-side Components

Manager with API Gateway

It routes external API requests from front-end services to individual agents. It also monitors and scales the cluster of multiple agents (a few tens to hundreds).

- Package namespace: `ai.backend.gateway` and `ai.backend.manager`
- <https://github.com/lablup/backend.ai-manager>

Not Logged In

[Login](#)
[Register](#)
[Lost Login?](#)
[Login with OpenID](#) 
[Login with Google](#) 

Status

[Nothing to report](#)



USER MANUALS

[API Overview](#)[Python Client Library](#)

API COMMON REFERENCE

[API and Document Conventions](#)[Authentication](#)[Rate Limiting](#)[JSON Object References](#)

USER API REFERENCE

[Introduction](#)[Kernel Management](#)[Code Execution \(Query Mode\)](#)[Code Execution and Monitoring
\(Streaming Mode\)](#)[Code Execution \(Batch Mode\)](#)[Virtual Folders](#)

Backend.AI Documentation

Latest API version: v3.20170615 (beta)

Backend.AI is a hassle-free backend for AI programming and service. It runs arbitrary user codes safely in resource-constrained environments, using Docker and our own sandbox wrapper.

Backend.AI supports various programming languages and runtimes, such as Python 2/3, R, PHP, C/C++, Java, Javascript, Julia, Octave, Haskell, Lua and NodeJS, as well as AI-oriented libraries such as TensorFlow, Keras, Caffe, and MXNet.

FAQ

vs. Notebooks

Product	Role	Problem and Solution
Apache Zeppelin, Databricks, etc.	Notebook-style document + code	Insecure host resource usage



USER MANUALS

API Overview

Python Client Library

API COMMON REFERENCE

API and Document Conventions

Authentication

Access Tokens and Secret Key

Common Structure of API Requests

Common Structure of API Responses

Signing API Requests

Example Requests and Responses

Rate Limiting

JSON Object References

USER API REFERENCE

Introduction



Authentication

Access Tokens and Secret Key

To make requests to the API server, a client needs to get a pair of an access token and a secret key as specified in `/gsg/registration`. The server uses access tokens to identify each client and secret keys to verify integrity of API requests as well as to authenticate clients.

Warning

For security reasons (to avoid exposition of your API access key and secret keys to arbitrary Internet users), we highly recommend to setup a server-side proxy to our API service if you are building a public-facing front-end service using Sorna.

For local deployments, you may create a master dummy pair in the configuration (TODO).

Common Structure of API Requests

오픈소스와 래블업

오픈소스가 준 것들





- 콘크리트화된 소스
 - 사용자 참여를 위한 문서화
 - 코드 커버리지 / CI 도입을 통한 코드 관리 체계화
- 새로운 기회
 - 오픈소스가 진입할 수 있는 시장
 - Backend.AI 가 진출할 수 있는 시장들
- 회사 가치
 - 말 대신 코드로 회사를 소개
 - '기술 회사' 인 것을 증명하는 법: 더 뚜렷한 과시 방법이 있을까요?
- 팀



- <https://github.com/lablup/backend.ai>
 - Backend.AI 개발에 참여하실 분 모십니다!
 - ✓ 삼도 같이 뜨면 낫겠죠? 😊
 - 2017년 10월 v1.0 릴리즈
 - ✓ (드디어 기다리고 기다리던) 설치 및 개발 매뉴얼 제공!
- 향후 로드맵
 - 스케줄러 기능 강화
 - ✓ Hybrid cloud 및 on-premise 연동
 - 오토스케일링 기능 강화
 - ✓ 장시간 연산 세션을 위한 scale-in protection
 - ✓ cpu/memory/gpu slot 여유 용량에 따라 cold/hot instance group 운영

마무리



THANK YOU

질문은 contact@lablup.com 으로!